# ON THE FINITE SAMPLE PERFORMANCE OF THE NEAREST NEIGHBOR CLASSIFIER*

Demetri Psaltis[†], Robert R. Snapp[‡],
and
Santosh S. Venkatesh[§]

## ABSTRACT

The finite sample performance of a nearest neighbor classifier is analyzed for a two-class pattern recognition problem. An exact integral expression is derived for the $m$-sample risk $R_m$ given that a reference $m$-sample of labeled points, drawn independently from Euclidean $n$-space according to a fixed probability distribution, is available to the classifier. For a family of smooth distributions characterized by asymptotic exapansions in general form, it is shown that the $m$-sample risk $R_m$ has a complete asymptotic series expansion $R_m \sim R_\infty + \sum_{k=1}^{\infty} c_k m^{-k/n}$ $(m \to \infty)$ where $R_\infty$ denotes the nearest neighbor risk in the infinite-sample limit. Improvements in convergence rate are shown under stronger smoothness assumptions, and in particular, $R_m = R_\infty + O(m^{-2/n})$ if the class-conditional probability densities have uniformly bounded third derivatives on their probability one support. This analysis thus provides further analytic validation of Bellman's curse of dimensionality. Numerical simulations corroborating the formal results are included, and extensions of the theory discussed. The analysis also contains a novel application of Laplace's asymptotic method of integration to a multidimensional integral where the integrand attains its maximum on a continuum of points.

*The work reported here was supported in part by the Air Force Office of Scientific Research under grant AFOSR 89–0523 to Santosh S. Venkatesh.

[†]Department of Electrical Engineering, California Institute of Technology, Pasadena, CA 91125; *electronic mail*: psaltis@sunoptics.caltech.edu

[‡]CS/EE Department, University of Vermont, Burlington, VT 05405; *electronic mail*: snapp@uvm.edu

[§]Department of Electrical Engineering, University of Pennsylvania, Philadelphia, PA 19104; *electronic mail*: venkatesh@ee.upenn.edu